Biophysical Journal Volume 98 April 2010 1247–1256

Predicting Gene-Regulation Functions: Lessons from Temperate Bacteriophages

Vladimir B. Teif*
Research Group Genome Organization & Function, Deutsches Krebsforschungszentrum and BioQuant, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany; and Institute of Bioorganic Chemistry, Belarus National Academy of Sciences, Kuprevich 5/2, 220141, Minsk, Belarus

ABSTRACT Gene-regulation functions (GRF) provide a unique characteristic of a cis-regulatory module (CRM), relating the concentrations of transcription factors (input) to the promoter activities (output). The challenge is to predict GRFs from the sequence. Here we systematically consider the lysogeny-lysis CRMs of different temperate bacteriophages such as the *Lactobacillus casei* phage A2, *Escherichia coli* phages λ, and 186 and *Lactococcal* phage TP901-1. This study allowed explaining a recent experimental puzzle on the role of Cro protein in the lambda switch. Several general conclusions have been drawn: 1), long-range interactions, multilayer assembly and DNA looping may lead to complex GRFs that cannot be described by linear functions of binding site occupancies; 2), in general, GRFs cannot be described by the Boolean logic, whereas a three-state non-Boolean logic suffices for the studied examples; 3), studied CRMs of the intact phages seemed to have a similar GRF topology (the number of plateaus and peaks corresponding to different expression regimes); we hypothesize that functionally equivalent CRMs might have topologically equivalent GRFs for a larger class of genetic systems; and 4) within a given GRF class, a set of mechanistic-to-mathematical transformations has been identified, which allows shaping the GRF before carrying out a system-level analysis.

INTRODUCTION

Gene regulation is currently studied from two points of view. Classical molecular biology is exploring the mechanistic details of transcription and translation (1,2), whereas the emerging field of systems biology tends to quantify gene expression on a larger scale without knowing physical details of macromolecular interactions (3–5). The need to bridge the two fields is evident after the first attempts to predict gene expression from the DNA sequence (3,4), which showed that a systematic relation of mechanistic protein-DNA and protein-protein interactions to the mathematical logic of gene regulatory elements remains lacking and required.

The simplified view of transcription factors acting at discrete predefined binding sites either as activators or repressors is obviously too far from the real world. On the other hand, connecting all-atom calculations with system-level analysis is not realistic at present. To solve this problem, we focus at the intermediate level. We consider, with single-nucleotide resolution, an elementary genetic module defined as the smallest group of *cis*-acting regulatory DNA regions, which may be mechanistically decoupled from the entire system and still retain its properties of regulating the given promoter(s). The expression of such a CRM may be determined as a unique function of the concentrations of transcription factors (TF), RNA polymerase (RNAP), and other protein players (Fig. 1). There have been several names for such functions in the literature, including the regulation factors (5), logic functions (6), input functions (7), *cis*-regulatory input functions (8,9), and gene-regulation functions (GRFs) (10). GRFs are uniquely determined by the sequence and epigenetic covalent modifications. Several strategies have been proposed to extract GRFs from expression data using the black box approach (6–9). Here, we try to predict GRFs from mechanistic molecular considerations. Recently, a first systematic sequence-to-function study has been carried out for genetic modules regulated by one transcription factor (11). We consider genetic modules regulated by two different transcription factors.

We consider several CRMs responsible for the lysogeny/lysis decision in temperate bacteriophages. Temperate bacteriophages have one common feature: when a phage infects a bacterium, it follows either a lytic or lysogenic pathway (2,12). In the lytic pathway, the genes responsible for the construction of viral capsids become activated. On the other hand, in the lysogenic pathway only the repressor protein CI is being produced, which silences the rest of the phage's genome integrated into the infected bacteria until external signals such as UV irradiation command a switch from the lysogeny to lysis. Since the first quantitative models of the lysogeny/lysis switch in *Escherichia coli* phage λ (13,14), much attention has been devoted to this particular phage (2) and a group of sequence-related, so-called lambdoid phages (12,15,16). Classical theoretical approaches in this field have been based on assigning a statistical weight for each combination of bound proteins and thus calculating the partition function (14,16–20). Another possibility is to operate on a

Submitted July 23, 2009, and accepted for publication November 25, 2009.
*Correspondence: Vladimir.Teif@bioquant.uni-heidelberg.de

Abbreviations used: bp, basepair; CRM, *cis*-regulatory module; GRF, gene-regulation function; RNAP, RNA polymerase; TF, transcription factors; UV, ultraviolet.

Editor: Andre Levchenko.

0006-3495/10/04/1247/10 $2.00
doi: 10.1016/j.bpj.2009.11.046

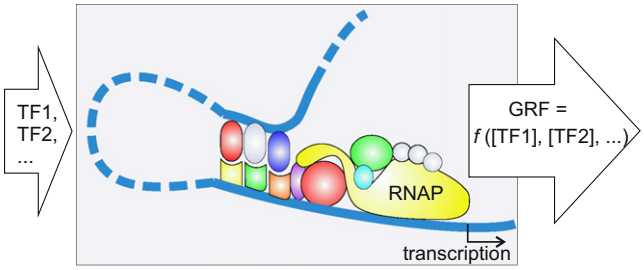

FIGURE 1 Schematic representation of a CRM and the corresponding GRF. Transcription starts after RNAP recruitment and activation, which may or may not require binding of multiple architectural proteins and transcription factors and DNA looping. The CRM is considered as a black box characterized by its GRF. The input variables are the concentrations of transcription factors. The output of the function is the promoter activity given, e.g., by the transcription rate, which depends on the RNAP binding probability, the rates of transcription initiation, elongation, and other kinetic parameters.

smaller scale, such as a nucleotide, and enumerate all states of each nucleotide, not the states of each protein. This increases the number of states, but decreases the number of parameters and explicit assumptions (21). For example, the early model of the bacteriophage λ lysogeny-lysis switch contained just eight states for the rearrangements of CI and Cro proteins at λ O_R (14) (Fig. 2). Later models deal with 40 states for CI/Cro/RNAP binding at λ O_R (19), or 64 states taking into account the O_R-O_L loop (20), whereas a single-nucleotide description requires 518 states for each basepair of the O_R-O_L contact region (21). A single-nucleotide description can be achieved in the frame of several methods described recently (4,11,22–23); one of such methods is the transfer matrix formalism, which is suitable as a systematic tool to study DNA-protein-drug binding in gene regulation (21), signal transduction on the membrane (25), and epigenetic regulation in chromatin (26). Here we have applied this method to calculate GRFs of the lysogeny-lysis switch elements of temperate bacteriophages shown in Fig. 2.

We start with a phage A2, which infects *Lactobacillus casei* and *Lactobacillus paracasei* strains of lactic acids bacteria (27,28). We construct and validate what we believe is a first mathematical model for this system. In a somewhat better studied *E. coli* phage 186, gene regulation of the lysogeny-lysis switch has been characterized previously by an intricate model very different from the conventional λ-models (15,16). Here we consider phage 186 in the frame of exactly the same mathematical formalism as other phages. The best-defined system remains phage λ, for which we consider two different mutations and provide a refinement to a previous model (21) after an improved model of promoter activation suggested recently (29). The calculations have clarified the roles of CI and Cro explaining the recent experimental puzzles in λ-mutants lacking Cro regulation. We end up with a general set of mechanistic-to-mathematical transformations for the studied class of two-variable GRFs, which seem to be essentially non-Boolean and nonlinear.

MATERIALS AND METHODS

The calculations have been carried out with the help of the transfer matrix formalism. Details of this method are provided in our previous publications (21,25,26) and in the Supporting Material. The idea of the method is to consider the DNA molecule as a lattice of N units (bp), each unit being characterized by a matrix of statistical weights corresponding to all its possible states. The weights are composed of the sequence- and context-specific DNA-protein binding constants, protein-protein interaction parameters and concentrations. The methodology consists of constructing transfer matrices for each basepair, applying the boundary conditions, and finally calculating protein binding maps as derivatives of the partition function, which is determined as a product of all transfer matrices (21).

There were only two variables in the calculations: the concentrations of free CI and Cro dimers. Unless specified otherwise, RNAP concentration was 30 nM consistent with previous estimates (18,30). RNAP length was set to $m = 35$ bp. CI and Cro lengths and other distances used in the calculations are indicated in Fig. 2. Long-range RNAP-RNAP interactions were characterized according to the experimental potential measured with single-nucleotide resolution up to the distances of 15 bp (31) as summarized in Table S4. Phage λ energetic parameters were taken as in a previous study (21) based on the experimental data (19,20,31,32). Phage A2, and 186 parameters were estimated from available experimental data as described in the text below. These parameters may be found in Table S3, Table S4, Table S5, and Table S6. Binding to all possible binding sites has been considered with single-nucleotide resolution, including nonspecific sites. In particular, when DNA looping was taken into account, all possible configurations of protein bridges at O_R have been considered with single-nucleotide resolution.

RESULTS

Phage A2 calculations

The regulation of the lysogeny-lysis switch module of phage A2 involves three protein players: CI, Cro, and RNAP. The activation of RNAP at A2 P_R and P_L promoters does not require additional contacts with CI or Cro because the promoters are strengthened by the UP sequences, which recognize the σ-RNAP subunit and facilitate the formation of transcriptionally active open complex (28). The parameters used in our modeling of phage A2 were extracted from the experimental data (28,33) as summarized in Table S3. CI and Cro dimers cover $m = 17$ bp on binding to the DNA. The nonspecific CI dissociation constant was $K(ns, \mathrm{CI}) = 100\ \mu$M (33); CI binding to the sites O_1 and O_2 was characterized by $K(O_1, \mathrm{CI}) = K(O_2, \mathrm{CI}) = 7$ nM based on the total CI binding curve (28); CI binding to O_3 was ~4 times weaker based on the reported concentration (28 nM) required to protect this site in the footprinting experiments (28). RNAP dissociation constants for the left and right promoter, $K(P_L, \mathrm{RNAP})$ and $K(P_R, \mathrm{RNAP})$, were chosen according to the experiments, which show that RNAP binding is three times stronger than CI binding, and O_R is 10 times stronger than O_L (28,34). Cro binding constants were based on the experimental value of 6 nM for the total Cro binding curve, and the ratio of concentrations 225:150:75 nM, which are required to protect the binding sites O_1, O_2, and O_3 in the footprinting experiments (34). According to the experiments, CI binding is cooperative ($w = 10$) whereas Cro is not (28,34). The RNAP-RNAP long-range interaction potential

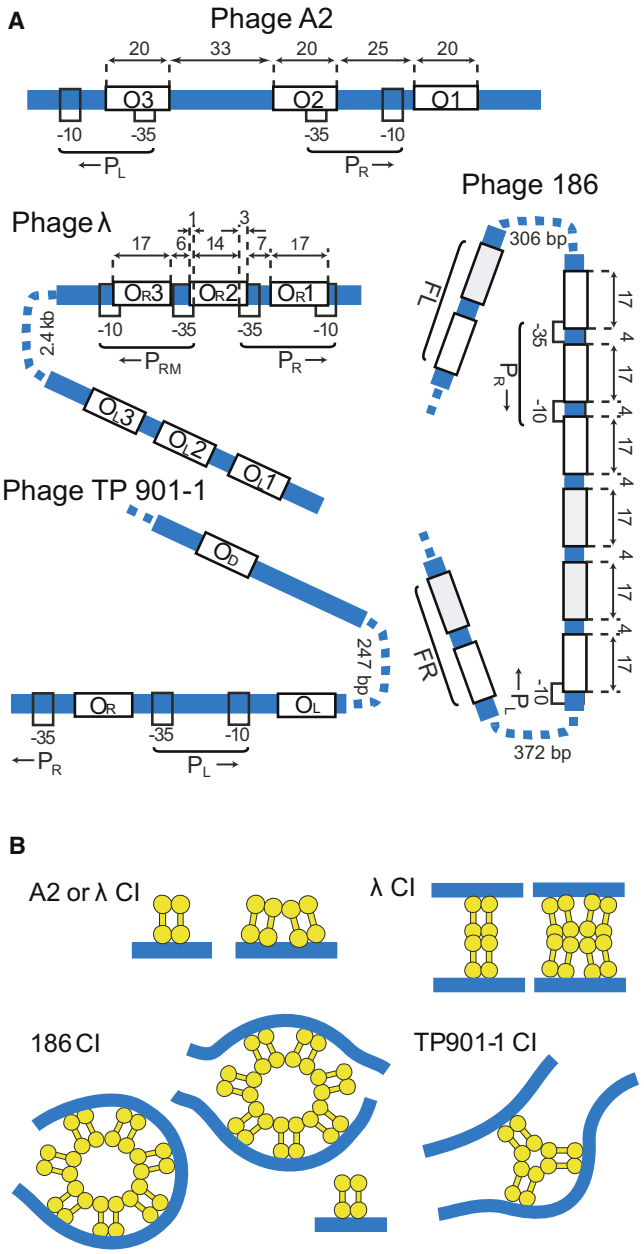

FIGURE 2 (*A*) Schemes of the lysogeny/lysis switch regions in phages A2, λ, 186, and TP 901-1. Phage A2: Three regulatory protein binding sites, O_1, O_2, and O_3 overlap with two back-to-back promoters P_L and P_R. Phage λ: O_R operator contains two back-to-back promoters P_{RM} and P_R, overlapping with the regulatory sites O_R1, O_R2, and O_R3. λ CI dimers can form tetramers and octamers to bridge O_R to a similarly organized O_L operator separated by ~2.4 kb (2). Phage 186: promoters P_R and P_L directed head-to-head are separated by 62 bp. 186 CI proteins can bind DNA as dimers and can also form wheels of seven CI dimers. Three strong CI binding sites overlap with P_R. One strong CI binding site overlaps with P_L, and two weaker sites lay between P_L and P_R. *FL* and *FR* regions separated by ~300 bp may be bridged to the P_L-P_R region by CI multimers (16). Phage TP 901-1: back-to-back promoters P_R and P_L are regulated by CI binding sites O_L and O_R separated by 63 bp. Binding of CI at O_L is expected to mediate steric hindrance for RNA polymerase interaction at the lytic P_L promoter. The O_D site located 247 bp from the O_L site may be bridged to the O_L-O_R region by a protein multimer consisting of three CI dimers (50). An additional binding site O_M overlapping with P_R has been proposed theoretically (51) (not shown in the figure). (*B*) Multimerization states of CI proteins in phages A2, λ, 186, and TP 901-1.

was taken as in the *E. coli* λ-model (21) based on the experimental promoter interference data (31) as summarized in Table S4. This model does not have any fitting parameters.

Fig. 3 shows the probability of RNAP binding to A2 P_R and P_L promoters calculated for different RNAP concentrations. The promoter activity is proportional to the probability of RNAP binding to the promoter. In Fig. 3 *A*, increasing concentration of CI dimers from nanomolar to micromolar range leads to the gradual increase in the expression of P_R in the absence of Cro. These calculations are quantitatively consistent with the experimental observation that the repression of P_R expression starts at [CI] = 7.4 nM, whereas P_R is completely repressed at [CI] = 180 nM (28).

Fig. 3, *B–D*, show the GRFs for A2 P_R and P_L operators calculated using the concentrations of CI and Cro dimers as variables. The working concentration of RNAP was 30 nM, corresponding to the available experimental data of A2 phage (28). Furthermore, in analogy with *E. coli* (18,30), we assume that 30 nM is the in vivo RNAP concentration in A2 hosts *L. casei* and *L. paracasei*, because they are approximately of the same size as *E. coli* (35). Therefore the GRFs in Fig. 3 *C* are closer to the reality than the GRFs calculated for other RNAP concentrations in Fig. 3, *B* and *D*. Fig. 3 *C* shows that the interval of CI and Cro concentrations exists where P_L is activated whereas P_R is not. The characteristic feature of A2 P_R and P_L GRFs is the existence of a flat plateau in the region of small concentrations and an increase of P_L expression coupled to the decrease of P_R expression in the intermediate concentration interval. At higher concentrations, P_L is again repressed.

Phage λ calculations: the RNAP recruitment model

λ CI and Cro bind DNA as dimers as well as in the case of phage A2. Cro-Cro interactions can be characterized by the standard McGhee-von Hippel cooperativity (36), whereas CI-CI interactions are more intricate. CI-CI interactions are pairwise and not symmetric: dimers of λ CI tend to form tetramers or octamers, but not trimers or pentamers (19). This means that two adjacent λ CI dimers interact cooperatively, whereas a third dimer has to find a new partner to form an interacting pair. This type of interaction requires a special model called *f*-mer assembly in the frame of the transfer matrix formalism (21) (here $f = 2$). RNAP bound at λ P_{RM} promoter requires the activation through a CI-RNAP contact, in contrast to phage A2 where a sole RNAP binding to P_L is enough to start transcription. Another feature of phage λ is the possibility of DNA looping between O_R and O_L bridged by CI multimers (Fig. 2 and Fig. S1) (2). In this case, there are two layers of potential protein binding sites, which should be treated by the multilayer matrix model (21) detailed in Fig. S1 and Table S2. This model provides

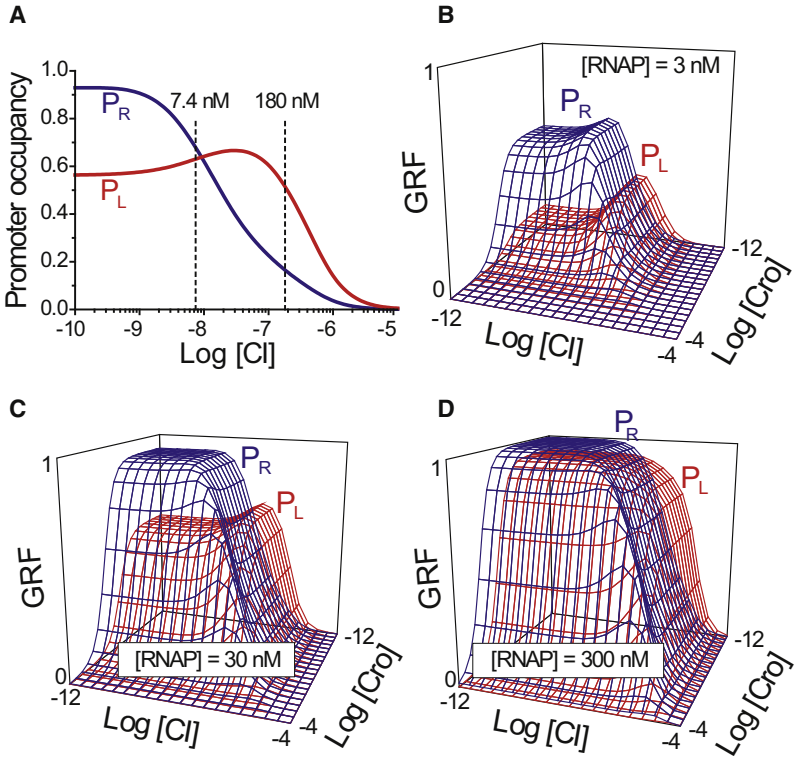

FIGURE 3 (*A*) RNAP occupancy at A2 P_R and P_L promoters as a function of CI concentration. Solid lines are calculated using the transfer matrix formalism. Dashed lines correspond to the experimental observations of the CI concentrations required to start the repression of P_R activity (7.4 nM) and to repress is completely (180 nM) (28). [RNAP] = 30 nM. (*B–D*) Phage A2 GRFs are set equal to the promoter occupancies, calculated as a function of [CI] and [Cro] at [RNAP] = (*B*) 3 nM, (*C*) 30 nM, and (*D*) 300 nM.

an accurate bookkeeping considering whether a given nucleotide is covered by a protein, whether this protein is further bound by another protein in the next protein layer, and whether a DNA-protein-protein-DNA bridge is created at a given position along the DNA.

We will consider below two experimentally available λ-mutations whose gene regulation at O_R is determined either by the RNAP recruitment or by the open complex stabilization mechanism. In the wild-type λ, the open complex stabilization mechanism is realized: CI-RNAP contact does not alter initial RNAP-DNA binding, but induces conformational changes in the RNAP-DNA complex at a later stage. This leads to the increase in the rate of a transcriptionally active open complex formation (37). Another λ-mutant has been constructed in the past, which does not affect the closed-open complex equilibrium but activates P_{RM} through RNAP recruitment, increasing RNAP binding to the DNA through cooperative RNAP-CI interactions (38). It was noted recently that the cooperativity through the open complex stabilization mechanism leads to significant system-level differences as compared to the energetically equivalent cooperativity through RNAP recruitment (29).

Fig. 4 *A* shows λ P_R and P_{RM} GRFs, calculated taking into account DNA-protein interactions and CI-CI and Cro-Cro cooperativity, but not taking into account CI-RNAP interactions and DNA looping. The GRFs in Fig. 4 *A* are identical to those calculated by Darling et al. (19).

Fig. 4 *B* adds to these interactions the CI-RNAP cooperativity in the frame of the RNAP recruitment mechanism. A cooperativity parameter $w = 10$ was estimated from the early experiments showing 10-fold P_{RM} activation by a CI-RNAP contact (32). It is noted that the P_{RM} GRF in Fig. 4 *B* has two plateaus instead of one in Fig. 4 *A*. A two-plateau function indicates that there are two regimes for this GRF, which reveals at the system level as the λ-switch.

The model from Fig. 4 *B* is further refined in Fig. 4 *C* where anticooperative interactions between RNA polymerases are taken into account (31). As seen from Fig. 4 *C*, the introduction of the experimental RNAP-RNAP interaction potential increases the vertical gap between P_R and P_{RM} plateaus at small CI concentrations.

Fig. 4 *D* shows the GRFs calculated for the model in Fig. 4 *B* taking into account the possibility of P_L-P_R looping. Fig. 4 *D* indicates that DNA looping confines the P_{RM} plateau by an almost vertical wall, thus prohibiting high CI concentrations. The picture shown in Fig. 4 *D* refined by RNAP-RNAP interactions as in Fig. 4 *C* would result in widening of the gap between the P_R and P_{RM} plateaus at small CI concentrations (data not shown). An additional difference between the lytic and lysogenic P_{RM} levels would arise if we account for possible multiple RNAP-CI contacts in the looped O_R-O_L state, which have been proposed recently (39).

Phage λ calculations: the open complex stabilization model

In the open complex stabilization model, initial RNAP binding to the DNA does not depend on the regulatory proteins. Instead, the RNAP-DNA complex undergoes a transition to the transcriptionally active state depending on the presence of CI at O_R2.

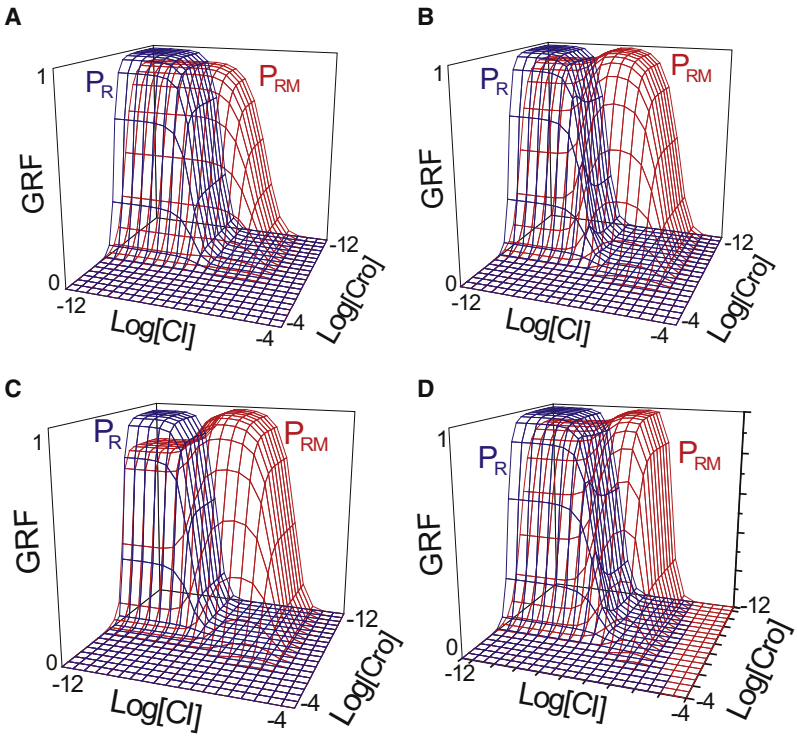

FIGURE 4 Phage λ GRFs calculated for the λ-mutant with the RNAP recruitment mechanism (38). GRFs are set equal to the promoter occupancies. (*A*) The model of Darling et al. (19) with cooperative interactions between CI and Cro proteins. (*B*) In addition to CI-CI and Cro-Cro interactions, RNAP-CI cooperativity is taken into account. (*C*) Anticooperative interactions between RNAPs bound to adjacent promoters are taken into account according to the experimental potential (31). (*D*) The O_R model from *B* is calculated taking into account the possibility of the O_R-O_L loop formation.

Fig. 5 *A* shows the probabilities of promoter occupancies by RNAP for this model, taking into account Cro-Cro, CI-CI, and RNAP-RNAP interactions as in Fig. 4 *C*. Here RNAP binds DNA noncooperatively. The probability of CI binding at O_R2 is shown in Fig. 5 *B*. Fig. 5 *C* shows GRFs calculated for the linear O_R operator in the absence of O_L. These GRFs now represent the activities (expression levels) of promoters P_R and P_{RM}, which are not necessarily proportional to the promoter occupancy. In particular, the activity of P_R, GRF(P_R) remains proportional to P_R occupancy by RNAP. On the other hand, P_{RM} activity denoted as GRF(P_{RM}) is a more complex function of both the promoter

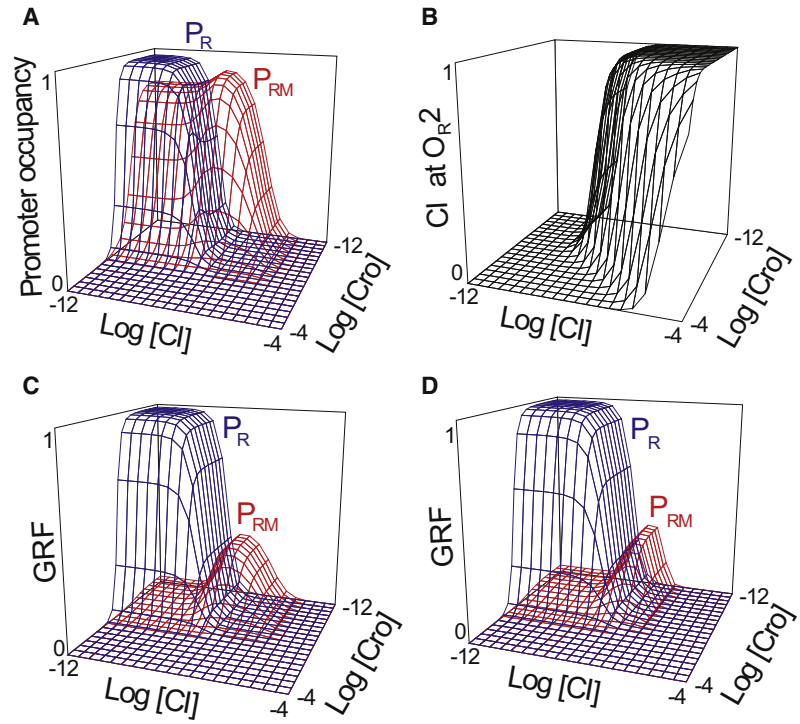

FIGURE 5 Phage λ GRFs calculated for the wild-type λ. RNAP-CI contact interactions affect the rate of the closed-open RNAP complex transition, but do not affect the initial RNAP binding. (*A*) RNAP occupancies at promoters. (*B*) CI occupancy at P_R2. (*C*) GRFs are set equal to the promoter activities in the absence of P_L-P_R looping given by Eq. 1. (*D*) GRFs are set equal to the promoter activities taking into account P_L-P_R looping given by Eq. 2.

occupancy by RNAP, $c(\text{RNAP}, P_{\text{RM}})$, and the occupancy of CI at $O_{\text{R}}2$, $c(\text{CI}, O_{\text{R}}2)$, as given below. Different transcription rates have been estimated for this system with CI bound or absent at $O_{\text{R}}2$ (20). This allows us to define the following linear dependence for the P_{RM} activity:

$$\text{GRF}(P_{\text{RM}}) = c(\text{RNAP}, P_{\text{RM}}) \times (k_{\text{b}} + k_{\text{a}} \times c(\text{CI}, O_{\text{R}}2)), \quad (1)$$

where k_{a} and k_{b} are experimental P_{RM} transcription rates normalized relatively to the P_{R} rate: $k_{\text{b}} = 0.04$ for the basal (nonactivated) P_{RM} state, $k_{\text{a}} = 0.35$ for activated P_{RM} with CI bound at $O_{\text{R}}2$ (20). The lytic interval of concentrations is characterized by a low plateau of GRF(P_{RM}).

Fig. 5 *D* shows GRFs calculated taking into account the O_{L}-O_{R} looping shown in Fig. 2 *C*. In this case, the GRF depends not only on the occupancy of the promoter by RNAP and the $O_{\text{R}}2$ site by CI, but also on the absence or presence of the O_{R}-O_{L} bridge, $c(\text{loop}, O_{\text{R}})$. A linear GRF dependence on these factors was chosen based on the experimental transcription rates as given by Eq. 2:

$$\begin{aligned}\text{GRF}(P_{\text{RM}}) = {} & c(\text{RNAP}, P_{\text{RM}}) \times (k_{\text{b}} + k_{\text{a}} \times c(\text{CI}, O_{\text{R}}2)) \\ & \times (1 - c(\text{loop}, O_{\text{R}})) + c(\text{RNAP}, P_{\text{RM}}) \\ & \times k_{\text{L}} \times c(\text{loop}, O_{\text{R}}),\end{aligned} \quad (2)$$

where, $k_{\text{L}} = 0.25$ is the looped state transcription rate of P_{RM} relatively to P_{R} (20).

Fig. 5 *D* shows that, similarly to the RNAP recruitment model in Fig. 4 *D*, DNA looping makes the shape of GRF(P_{RM}) more defined. A vertical wall confines the GRF(P_{RM}) peak and makes it narrower.

Phage 186 calculations

The regulation of phage 186 is even more intricate, involving two additional elements, *FR* and *FL*, which can bind CI multimers and assist DNA loop formation (15,16) (Fig. 2). Phage 186 CI proteins can form wheels containing seven CI dimers (15,16). The P_{L}-P_{R} region can be coiled around a CI wheel, or connected by a CI wheel to *FR* or *FL* regions. The lattice models constructed to describe this binding are described in details in the Supporting Material.

Fig. 6 shows GRFs calculated for the lysogeny-lysis CRM of phage 186 shown in Fig. 2. Phage-specific parameters were chosen based on the published experimental data (15,16). According to these data, three CI binding sites of length $m = 17$ bp overlapping with P_{R} and one site overlapping with P_{L} were defined as "strong" ($K = 0.57 \times 10^9\ \text{M}^{-1}$). The intermediate weak sites grayed out in Fig. 2 were characterized by a three-fold lower binding constant for CI dimers. The nonspecific binding was characterized in the same way as in the case of phage λ. RNAP binding was characterized by a binding constant equal to $10^{10}\ \text{M}^{-1}$ at P_{R}, and $7.2 \times 10^9\ \text{M}^{-1}$ at P_{L}. Nonspecific DNA binding

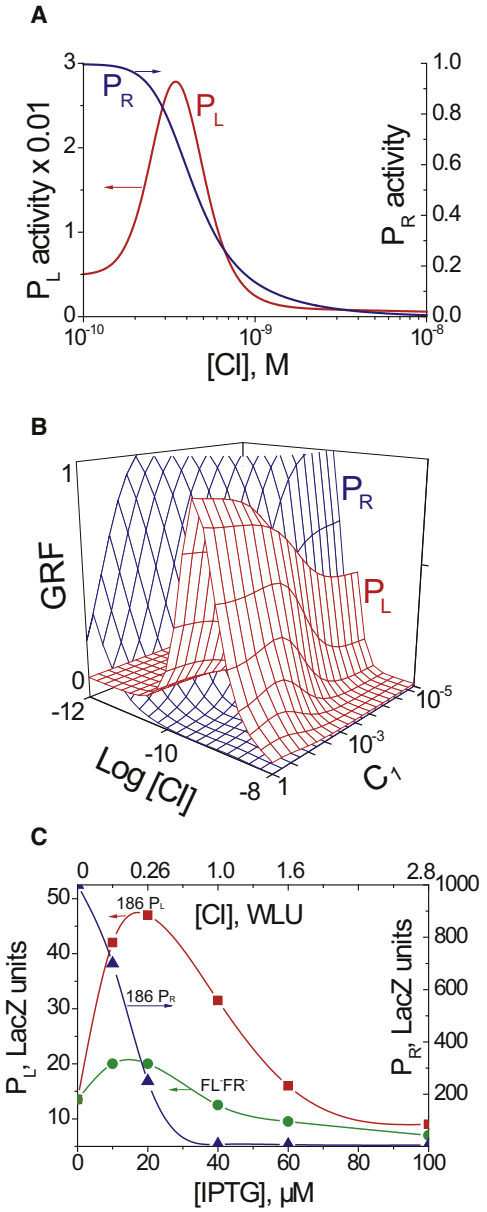

FIGURE 6 Phage 186 GRFs. (*A*) Activities of P_{R} and P_{L} promoters calculated for the 150 bp P_{R}-P_{L} region in the absence of *FL* and *FR* sites. (*B*) Activities of the same promoters calculated taking into account the *FR* and *FL* regions. Larger C_1 corresponds to a larger effect of *FR* and *FL* through DNA looping. (*C*) Experimental activities reconstructed from the published data (15): 1, 2, P_{R} and P_{L} activities of wild-type phage 186; 3, $FL^{-}FR^{-}$ phage 186 mutant without *FL* and *FR* regions.

constants were taken as in the phage λ system (Table S3, Table S4, Table S5, and Table S6). CI-CI interactions in the linear assemblies and inside the wheels were characterized by the cooperativity parameter $w = 30$. The 7-mers of CI dimers joining the regions separated by large distances along the DNA were characterized by a single protein type at a concentration $C_1 \times c_0(\text{CI})$, where $c_0(\text{CI})$ is the concentration of free CI dimers, $0 < C_1 < 1$. The parameter C_1 was varied in the calculations, studying the effect of external DNA

segments on the P_R-P_L region. The binding of RNAP adjacent to a CI multimer was characterized by the cooperativity parameter $w = 78$ (16). In addition, it was taken into account that a CI multimer bound close to RNAP activates P_L 108-fold by increasing its close-to-open complex formation rate (16). The latter resulted in Eq. 3 for phage 186 P_L GRF denoted as GRF(P_L):

$$\mathrm{GRF}(P_L) = C_2 \times c(\mathrm{RNAP}, P_L) \times (1 + 108^* c(\mathrm{CI}, P_L)). \quad (3)$$

Here c(RNAP, P_L) is the RNAP occupancy at P_L, c(CI, P_L) is the probability to find CI on the DNA near RNAP at P_L. The constant C_2 scales GRF(P_L) with respect to GRF(P_R) taking into account that P_L has a 12.4-fold lower closed-open complex formation rate, and a 15-fold lower elongation rate relative to P_R (16). GRF(P_R) was set equal to P_R occupancy by RNAP.

Fig. 6 *A* shows GRF(P_L) and GRF(P_R) calculated for the phage 186 FL^-FR^- mutant without *FL* and *FR* regions. These GRFs follow a trend observed previously for phages A2 and λ. GRF(P_R) is higher than GRF(P_L). Moreover, GRF(P_L) is confined by steep walls on both sides of small and large CI concentrations. Fig. 6 *B* shows GRF(P_L) and GRF(P_R) calculated for different values of C_1, reflecting the effect of *FL*-P_L and *FL*-P_R loops on the local CI concentration near P_L-P_R region. $C_1 \to 0$ means the absence of the effect of *FL* and *FR*, as in the FL^-FR^- mutant. This figure shows that additional recognition sequences at *FL* and *FR* help shaping the P_L logic function and make its maximum higher. This prediction is consistent with the experimental observations shown in Fig. 6 *C* (15,16). The changes in parameters C_1 and C_2 do not affect the GRF topology.

DISCUSSION

On the role of Cro in the lysogeny/lysis switch

A classical point of view is that the λ-switch is determined by the competition of Cro and CI proteins. When *E. coli* SOS system activated by UV radiation decreases CI concentration below a threshold level, Cro takes over and determines the switch to lysis (2). However, recent studies have reopened the question of the Cro role, suggesting that it might be not required to switch to lysis (40). To check for this possibility, two types of experiments have been carried out. In the first study, a module-replacement approach was used to substitute *cro* by *lac*. Lac protein is not naturally encoded in the λ-genome and is not interfering with CI binding at O_R (41). In the second approach, Cro binding sites at O_R have been mutated so that almost completely suppress its specific binding to this region (42). Both works have generated lambda mutants different from the native phage lambda, but they both argued that obtained mutants are reasonably comparable with lambda in many aspects but the absence of the Cro dependence. In both cases, it was shown that

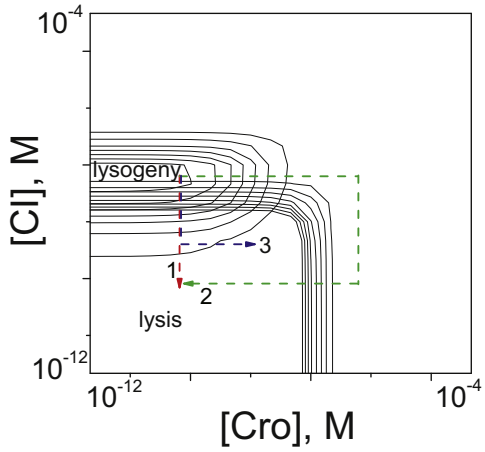

FIGURE 7 Phase diagram of the λ-switch plotted as a *Z*-projection of P_R and P_{RM} GRFs shown in Fig. 5 *D*. Multiple paths from the lysogeny to lysis are possible. Example paths are indicated by the arrows.

mutants lacking Cro are less effective to switch from the lysogeny to lysis. However, both works have shown that a small subpopulation of phages exists, which can switch to lysis without Cro. The authors of the two works have suggested confronting interpretations of these mutually consistent results, claiming that Cro is either important (42) or not important (41) for the switch. A recent theoretical work has failed to explain the results of the module-replacement experiments (41), concluding that the λ-mutants lacking Cro are not described by the standard λ-model (43). Although in the absence of thermodynamic data we can not construct GRFs for these λ-mutants here, we can still explain the puzzling role of Cro in the native λ.

Fig. 7 shows the phase diagram of the λ-switch plotted as a *Z*-projection of the GRFs from Fig. 5 *D*. The lines on the plot join the points with the same levels of P_R or P_{RM} expression, which results in two well-defined clouds reflecting two distinct regimes of functioning of the O_R operator. Each phage has its own fate and correspondingly belongs to the lysogenic or lytic regions of CI and Cro concentrations indicated in the figure. Phages may have different paths to transfer between the two regions of concentrations. These paths may depend on the history: whether it is the initial infection, induction of a prophage or some intermediate process. Two example paths are indicated by the arrows. The most straightforward path indicated by a label 1 is along the CI concentration axis. This path would not require changes in Cro concentrations, which explains why a certain subpopulation of phages can start prophage induction without Cro. On the other hand, phages with defective Cro regulation would not be able to use other paths such as the path indicated as 2 in Fig. 7. Experiments cited above suggest that not only the shortest path is realized in vivo; other paths are also important. In particular, one can hypothesize, that in some of the cells undergoing UV-irradiation, RecA proteins destroy most but not all CI dimers. In such cells,

Cro binding to O_R becomes necessary to prevent CI access to the DNA so that CI does not reestablish its lysogenic concentration, as schematically indicated by the path 3 in Fig. 7. Thus, the role of CI and Cro is different: all paths require changes in CI concentration, whereas some of them are independent of Cro. It should be noted, that Fig. 3 also predicts a similar role of Cro in phage A2, which might be verified by future experiments.

Gene regulation functions are not linear

A practical assumption used in many systems biology approaches is that the GRF can be represented by a linear algebraic function of the occupancies of binding sites of the CRM (4). In our calculations, a *cis*-regulatory module of phage A2 was indeed characterized by a GRF proportional to the promoter occupancy, which was a linear combination of the occupancies of other regulatory sites (Fig. 3). However, that was not the case when long-range DNA looping was taken into account for phage λ (Figs. 4 and 5). In this case, GRF(P_{RM}) was still proportional to the RNAP occupancy, but the dependence on CI occupancies was not linear. Note that no nonlinearity was assumed for the elementary reaction rates (although they might in principle change in the condensed DNA phase (44)). The nonlinearity of GRF(P_{RM}) comes in our calculations from its dependence on the looping probability, which is a nonlinear function of the reactant concentrations and the loop length (45,46). This explains how, starting from a linear GRF dependence on the reaction rates, we finally arrived to a nonlinear dependence of the GRF on the binding site occupancies as shown in Fig. 5. The correction was not very large in the λ case, but it might be larger for eukaryotic regulatory elements, which are much longer.

Gene-regulation functions are not Boolean

Many researchers believe that gene-regulation functions can be described by Boolean operators in analogy with computers. According to the definition, Boolean logic is based on two states: true and false (1 and 0 in computers). Any logic based on another number of states (e.g., three states: true, false, undefined) is not Boolean. In our case, there are two independent variables ([CI] and [Cro]) and one dependent variable (GRF). In the frame of the Boolean logic, each of these three variables can take two values. In total this gives rise to $2^{2\wedge 2} = 2^4 = 16$ possible combinations. The 3D plots corresponding to all 16 possible Boolean functions of two independent variables are shown in Fig. S2 (a similar graphical representation was provided in Fig. 1 of Mayo et al. (8)). A comparison of our calculated GRFs in Fig. 5 *B* with 16 Boolean operators in Fig. S2 shows that λ GRF(P_R) can be represented by the Boolean operators: GRF(P_R) = NOT ([CI] AND [Cro]). However, λ GRF(P_{RM}) cannot be obtained by means of any of 16 combinations of two-variable Boolean operators. Previously, it has been shown that the Lac operator of *E. coli* can be characterized by a combination of Boolean functions (8) whereas there are difficulties to interpret several other *E. coli* operators in this way (7). A similar controversial situation has been reported for *Drosophila* (47). Obviously, if at least one known GRF is not Boolean, the statement that all GRFs are Boolean is false. Thus we believe our work proves that GRFs are not Boolean functions of TF concentrations in a general case.

An example of a non-Boolean three-state logic, which is able to describe both λ GRF(P_R) and GRF(P_{RM}) is shown in Fig. S3. In a three-state logic GRFs can be described by $3^{3\wedge f}$ distinguishable functions, where *f* is the number of transcription factors (19,683 functions if there are two TFs). Here TF concentrations and GRF are described for example as small, medium, and large. In an intermediate 2.5-state logic where GRF is described only by two states (true, false) and TF concentrations by three states (small, medium, large), the number of distinguishable functions would be $2^{3\wedge f}$ ($2^9 = 512$ if there are two TFs). Note that both the 3-state and 2.5-state logic in Fig. S3 are not Boolean.

Functionally equivalent CRMs have topologically equivalent GRFs?

A series of in vitro experiments reported previously allowed our construction of gene-regulation functions for the lysogeny/lysis switch modules of *L. casei* phage A2 and *E. coli* phages λ and 186. These CRMs have been characterized by 3D GRF plots showing the promoter activity as a function of concentrations of CI and Cro dimers (Figs. 3–6). The mechanistic details of these systems are very different (Fig. 2), but the topology (the number of plateaus and peaks) of the GRFs of the intact phages appeared to be quite similar. This finding is even more striking because the GRFs of the two λ-variants in Figs. 4 and 5 differ stronger than the GRFs of the intact λ, A2, and 186.

Why the GRFs of the studied systems are similar? First, let us remind that the selection of genetic regions used in this study was according to their well-known biological function: the lysogeny-lysis switch in the corresponding phages. Thus one could argue that the mathematical similarity is due to their belonging to a single class of biological toggle-switchers. As an example of a different biological class of gene-regulation functions, one could consider *E. coli* Lac operator. In a large region of concentrations, Lac GRF has an almost linear dependence on the input concentration of cAMP (9), which resembles the behavior of an attenuator, unlike the toggle-switch behavior in our calculations. A possible experimental test for this possibility could be the determination of GRF for the lysogeny-lysis CRM in phage TP901-1 shown in Fig. 2.

Another possibility to explain the mathematical similarity of GRFs of the lysogeny-lysis switch modules of intact phages λ, A2, and 186 might be purely mechanistic: These

CRMs all have physically adjacent promoters regulated by overlapping binding sites of two TFs. If this explanation is true, then it might have important implications for mammalian organisms, where ~10% of genes are arranged head-to-head and characterized by bidirectional promoters in a physical proximity from each other (48).

Mechanistic rules to determine the GRF shape

The analysis of 3D GRFs plotted in our study has shown the following rules of the mechanistic-to-mathematical conversions that hold for the GRF class studied above:

1. Contact interactions between transcription factors make the GRF steeper.
2. The GRF shape becomes more defined in terms of plateaus and peaks in the working interval of RNAP concentrations (Fig. 3). GRFs calculated beyond in vivo RNAP concentrations (Fig. 3 *D*) do not have a distinct border between the expression regimes.
3. Anticooperative interactions between RNA polymerases bound to adjacent DNA sites increase the vertical gap between the promoter regimes, making a strong promoter stronger and a weak promoter weaker (Fig. 4 *C*).
4. A promoter activated by the open complex stabilization mechanism is regulated by a function with a narrow peak (Fig. 5) as compared to a multiplateau function of the promoter regulated by the RNAP recruitment mechanism (Fig. 4). This explains why large differences have been observed for these two energetically equivalent models at the system level (29).
5. Long-distance DNA looping by regulatory proteins can change the height of the plateau through additional activating/repressing protein-RNAP contacts, as well as to decrease the width of the plateau through a steric exclusion of RNAP (Figs. 4 *D* and 5 *D*). The latter effect protects the cell from the overexpression of regulatory proteins. The DNA loops in the studied toggle switch elements make the response to small changes in CI concentration sharper. This is different from a suggested role of looping in the Lac operator of *E. coli* where it makes the response smoother (17). This difference might be because the Lac operator belongs to the attenuator and not the toggle-switch class.

CONCLUSIONS

This work provides what we believe to be a conceptually new paradigm for predicting gene-regulation functions. The CRM is considered as the smallest region that can be decoupled from the entire system and still retain its function. Therefore it is impossible to isolate a noninteracting functional submodule inside a CRM. Any nucleotide change or change in the distances between specific sites would affect its GRF. That is why, for example, the λ-lac mutants of Atsumi and Little (41) could not be described by the standard λ-model in the calculations of Werner and Aurell (43). The approach presented here allows calculating this, but only if the sequence-affinity data are provided as was the case for native phages A2, λ, and 186 in Figs. 3–6. Obviously this method is more difficult to implement in comparison with the Boolean or linear functions. However, it is worth the efforts, because Boolean and linear approximations fail even for a simple λ-system as shown above. The GRFs calculated as in Figs. 3–6 allow determining the topology of the output signal but not the exact values. Further refinement of the parameters may be achieved on integrating GRFs into the system-level description and solving the differential equations at the system level. Recently, it was noted that parameter refinement is a very challenging task requiring special mathematical tricks even for simplified models treating TFs as activators and repressors (49). Our work suggests that physical constrains arising from mechanistic molecular interactions can accelerate the convergence and provide biologically meaningful solutions.

SUPPORTING MATERIAL

A list of notations, calculation algorithms six tables, three figures, and supplemental references are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(09)06093-7.

I thank Terence Hwa, Karsten Rippe, Ralf Metzler, and anonymous reviewers for valuable discussions or comments on the manuscript. I am grateful to the hospitality of the University of California, San Diego, Center for Theoretical Biological Physics, sponsored by National Science Foundation grant PHY-0822283, where this work was initiated. The support of the fellowship of the German CellNetworks Cluster of Excellence (EXC81) is greatly acknowledged.

REFERENCES

1. von Hippel, P. H. 2007. From "simple" DNA-protein interactions to the macromolecular machines of gene expression. *Annu. Rev. Biophys. Biomol. Struct.* 36:79–105.
2. Ptashne, M. 2004. Genetic Switch: Phage Lambda Revisited. Cold Spring Harbor Laboratory Press, New York.
3. Beer, M. A., and S. Tavazoie. 2004. Predicting gene expression from sequence. *Cell.* 117:185–198.
4. Segal, E., T. Raveh-Sadka, …, U. Gaul. 2008. Predicting expression patterns from regulatory sequence in *Drosophila* segmentation. *Nature.* 451:535–540.
5. Bintu, L., N. E. Buchler, …, R. Phillips. 2005. Transcriptional regulation by the numbers: models. *Curr. Opin. Genet. Dev.* 15:116–124.
6. Istrail, S., and E. H. Davidson. 2005. Logic functions of the genomic *cis*-regulatory code. *Proc. Natl. Acad. Sci. USA.* 102:4954–4959.
7. Kaplan, S., A. Bren, …, U. Alon. 2008. Diverse two-dimensional input functions control bacterial sugar genes. *Mol. Cell.* 29:786–792.
8. Mayo, A. E., Y. Setty, …, U. Alon. 2006. Plasticity of the *cis*-regulatory input function of a gene. *PLoS Biol.* 4:e45.
9. Setty, Y., A. E. Mayo, …, U. Alon. 2003. Detailed map of a *cis*-regulatory input function. *Proc. Natl. Acad. Sci. USA.* 100: 7702–7707.
10. Kim, H. D., and E. K. O'Shea. 2008. A quantitative model of transcription factor-activated gene expression. *Nat. Struct. Mol. Biol.* 15: 1192–1198.

11. Pérez, A. G., V. E. Angarica, …, A. T. Vasconcelos. 2009. From sequence to dynamics: the effects of transcription factor and polymerase concentration changes on activated and repressed promoters. *BMC Mol. Biol.* 10:92.
12. Abedon, S. T., and R. L. Calendar. 2005. The Bacteriophages. Oxford University Press, New York.
13. Kananyan, G. Kh., V. A. Ratner, and R. N. Tchuraev. 1981. Enlarged model of lambda phage ontogenesis. *J. Theor. Biol.* 88:393–407.
14. Ackers, G. K., A. D. Johnson, and M. A. Shea. 1982. Quantitative model for gene regulation by lambda phage repressor. *Proc. Natl. Acad. Sci. USA.* 79:1129–1133.
15. Dodd, I. B., and J. B. Egan. 2002. Action at a distance in CI repressor regulation of the bacteriophage 186 genetic switch. *Mol. Microbiol.* 45:697–710.
16. Dodd, I. B., K. E. Shearwin, and K. Sneppen. 2007. Modelling transcriptional interference and DNA looping in gene regulation. *J. Mol. Biol.* 369:1200–1213.
17. Saiz, L., and J. M. Vilar. 2006. DNA looping: the consequences and its control. *Curr. Opin. Struct. Biol.* 16:344–350.
18. Santillán, M., and M. C. Mackey. 2004. Why the lysogenic state of phage lambda is so stable: a mathematical modeling approach. *Biophys. J.* 86:75–84.
19. Darling, P. J., J. M. Holt, and G. K. Ackers. 2000. Coupled energetics of lambda Cro repressor self-assembly and site-specific DNA operator binding II: cooperative interactions of Cro dimers. *J. Mol. Biol.* 302:625–638.
20. Dodd, I. B., K. E. Shearwin, …, J. B. Egan. 2004. Cooperativity in long-range gene regulation by the lambda CI repressor. *Genes Dev.* 18:344–354.
21. Teif, V. B. 2007. General transfer matrix formalism to calculate DNA-protein-drug binding in gene regulation: application to O_R operator of phage lambda. *Nucleic Acids Res.* 35:e80.
22. Nechipurenko, Y. D., B. Jovanovic, …, G. V. Gursky. 2005. Quantitative methods of analysis of footprinting diagrams for the complexes formed by a ligand with a DNA fragment of known sequence. *Ann. N. Y. Acad. Sci.* 1048:206–214.
23. Wasson, T., and A. J. Hartemink. 2009. An ensemble model of competitive multi-factor binding of the genome. *Genome Res.* 19:2101–2112.
24. Reference deleted in proof.
25. Teif, V. B., D. Harries, …, A. Ben-Shaul. 2008. Matrix formalism for site-specific binding of unstructured proteins to multicomponent lipid membranes. *J. Pept. Sci.* 14:368–373.
26. Teif, V. B., and K. Rippe. 2009. Predicting nucleosome positions on the DNA: combining intrinsic affinities and remodeler activities. *Nucleic Acids Res.* 37:5641–5655.
27. Ladero, V., P. García, …, J. E. Suárez. 2002. Interaction of the Cro repressor with the lysis/lysogeny switch of the *Lactobacillus casei* temperate bacteriophage A2. *J. Gen. Virol.* 83:2891–2895.
28. García, P., V. Ladero, …, J. E. Suárez. 1999. Cooperative interaction of CI protein regulates lysogeny of *Lactobacillus casei* by bacteriophage A2. *J. Virol.* 73:3920–3929.
29. Gedeon, T., K. Mischaikow, …, E. Traldi. 2008. Binding cooperativity in phage lambda is not sufficient to produce an effective switch. *Biophys. J.* 94:3384–3392.
30. Arkin, A., J. Ross, and H. H. McAdams. 1998. Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected *Escherichia coli* cells. *Genetics.* 149:1633–1648.
31. Strainic, Jr., M. G., J. J. Sullivan, …, P. L. deHaseth. 2000. Promoter interference in a bacteriophage lambda control region: effects of a range of interpromoter distances. *J. Bacteriol.* 182:216–220.
32. Meyer, B. J., R. Maurer, and M. Ptashne. 1980. Gene regulation at the right operator (O_R) of bacteriophage λ. II. O_R1, O_R2, and O_R3: their roles in mediating the effects of repressor and Cro. *J. Mol. Biol.* 139:163–194.
33. Bakk, A., and R. Metzler. 2004. In vivo non-specific binding of λ CI and Cro repressors is significant. *FEBS Lett.* 563:66–68.
34. Ladero, V., P. García, …, J. E. Suárez. 1999. A2 cro, the lysogenic cycle repressor, specifically binds to the genetic switch region of *Lactobacillus casei* bacteriophage A2. *Virology.* 262:220–229.
35. Atanassova, M. R., V. Chipeva, …, T. Haertlé. 2001. Determination of the growth stages of *Lactobacillus paracasei subsp. paracasei* M3 from Bulgarian yellow cheese by electroconductivity. *J. Microbiol. Methods.* 46:227–233.
36. McGhee, J. D., and P. H. von Hippel. 1974. Theoretical aspects of DNA-protein interactions: cooperative and non-cooperative binding of large ligands to a one-dimensional homogeneous lattice. *J. Mol. Biol.* 86:469–489.
37. Nickels, B. E., S. L. Dove, …, A. Hochschild. 2002. Protein-protein and protein-DNA interactions of σ70 region 4 involved in transcription activation by λcI. *J. Mol. Biol.* 324:17–34.
38. Li, M., W. R. McClure, and M. M. Susskind. 1997. Changing the mechanism of transcriptional activation by phage lambda repressor. *Proc. Natl. Acad. Sci. USA.* 94:3691–3696.
39. Anderson, L. M., and H. Yang. 2008. DNA looping can enhance lysogenic CI transcription in phage lambda. *Proc. Natl. Acad. Sci. USA.* 105:5827–5832.
40. Svenningsen, S. L., N. Costantino, …, S. Adhya. 2005. On the role of Cro in lambda prophage induction. *Proc. Natl. Acad. Sci. USA.* 102: 4465–4469.
41. Atsumi, S., and J. W. Little. 2006. Role of the lytic repressor in prophage induction of phage λ as analyzed by a module-replacement approach. *Proc. Natl. Acad. Sci. USA.* 103:4558–4563.
42. Schubert, R. A., I. B. Dodd, …, K. E. Shearwin. 2007. Cro's role in the CI Cro bistable switch is critical for λ's transition from lysogeny to lytic development. *Genes Dev.* 21:2461–2472.
43. Werner, M., and E. Aurell. 2009. A computational study of lambda-lac mutants. *Phys. Biol.* 6:46007.
44. Teif, V. B. 2005. Ligand-induced DNA condensation: choosing the model. *Biophys. J.* 89:2574–2587.
45. Zurla, C., C. Manzo, …, L. Finzi. 2009. Direct demonstration and quantification of long-range DNA looping by the λ bacteriophage repressor. *Nucleic Acids Res.* 37:2789–2795.
46. van den Broek, B., M. A. Lomholt, …, G. J. Wuite. 2008. How DNA coiling enhances target localization by proteins. *Proc. Natl. Acad. Sci. USA.* 105:15738–15742.
47. Payankaulam, S., and D. N. Arnosti. 2008. Gene regulation: boundaries within limits. *Curr. Biol.* 18:R653–R655.
48. Lin, J. M., P. J. Collins, …, Z. Weng. 2007. Transcription factor binding and modified histones in human bidirectional promoters. *Genome Res.* 17:818–827.
49. Bauer, D. C., and T. L. Bailey. 2009. Optimizing static thermodynamic models of transcriptional regulation. *Bioinformatics.* 25:1640–1646.
50. Pedersen, M., and K. Hammer. 2008. The role of MOR and the CI operator sites on the genetic switch of the temperate bacteriophage TP901-1. *J. Mol. Biol.* 384:577–589.
51. Nakanishi, H., M. Pedersen, …, K. Sneppen. 2009. Modeling of the genetic switch of bacteriophage TP901-1: a heteromer of CI and MOR ensures robust bistability. *J. Mol. Biol.* 394:15–28.